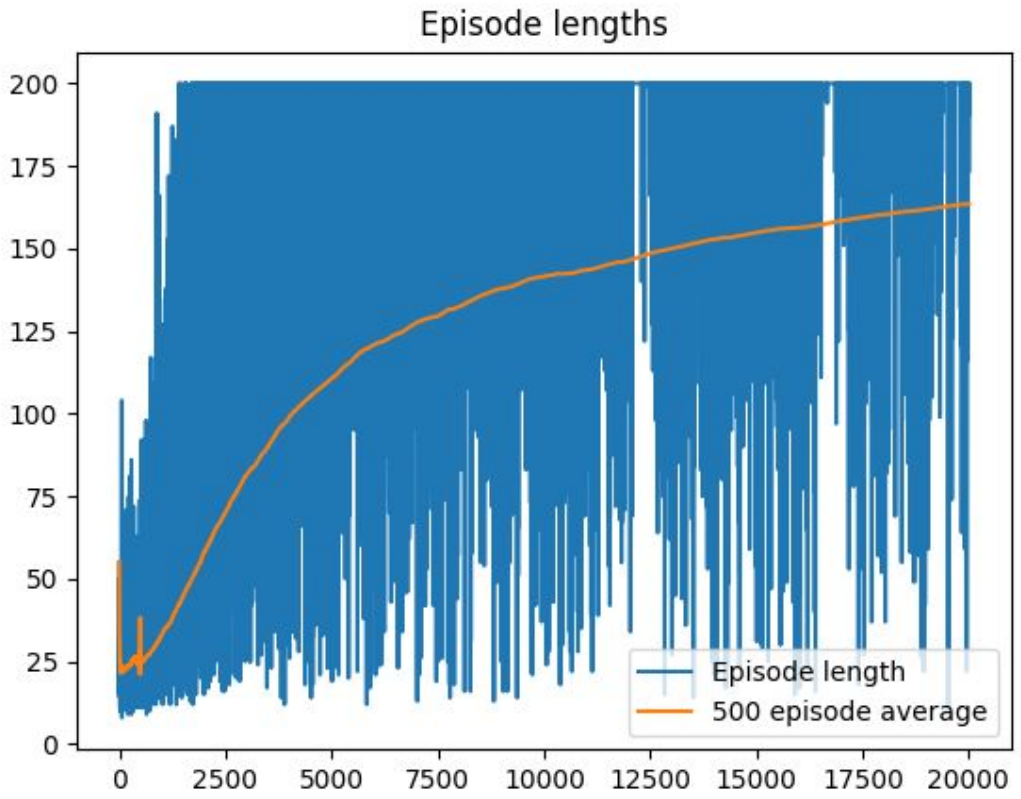
Reinforcement Learning
Exercise 3

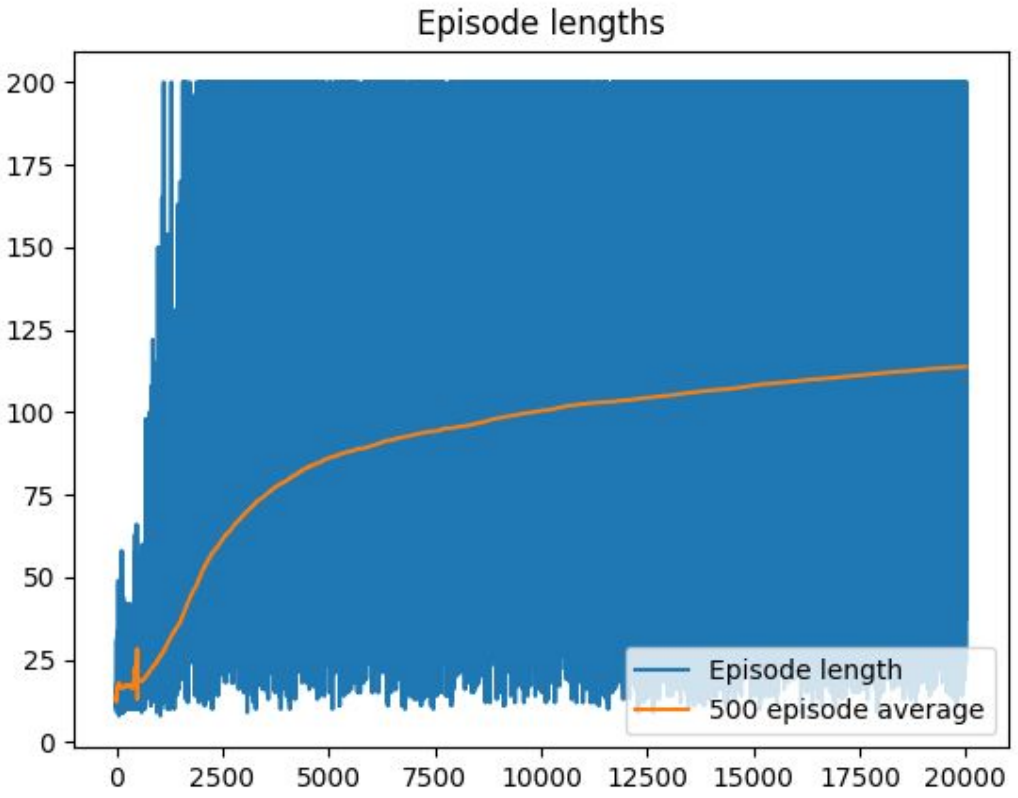
1 Q-Learning

1.1 Cartpole - Task 1

Exercise solved in file `qlearning.py`



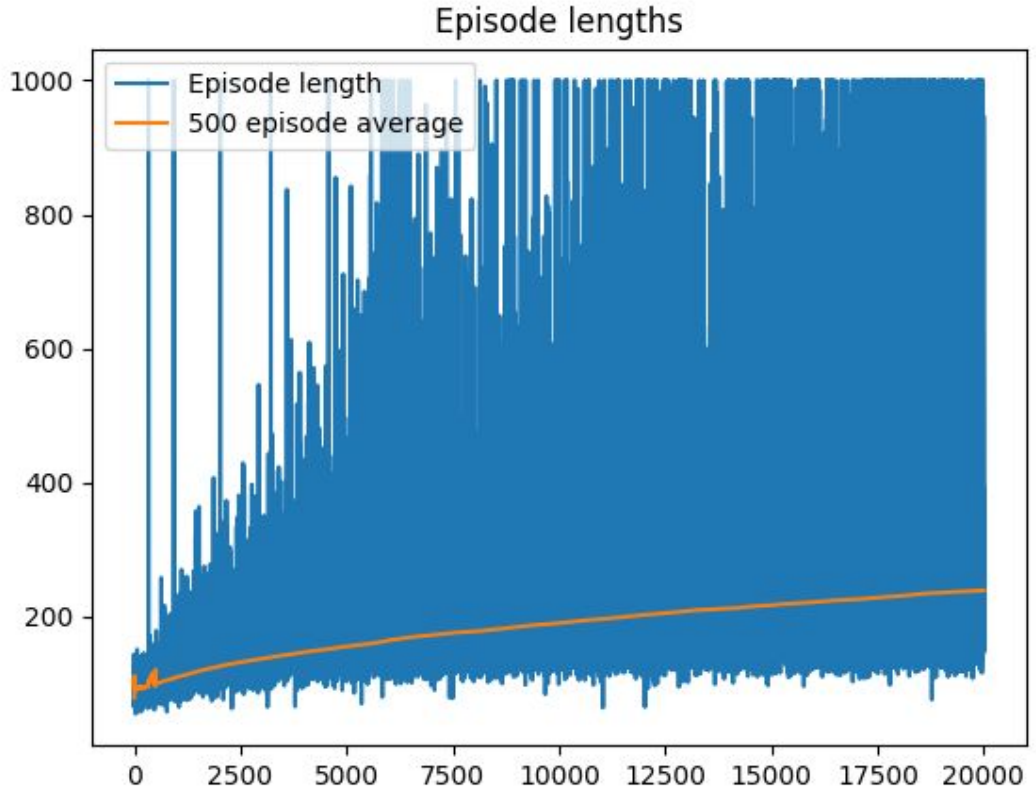
Plot of learning using epsilon greedy with infinite exploration.



Plot of learning using epsilon fixed to 0.4.

1.2 Lunar lander - Task 2

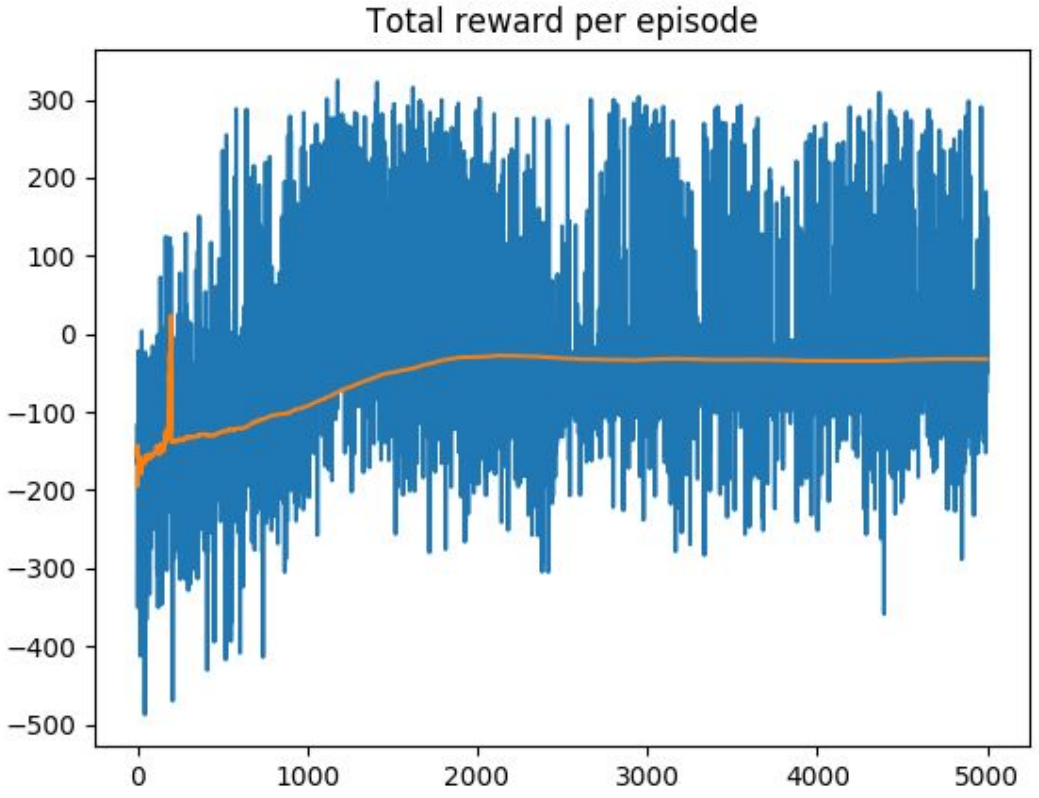
The lander learns to reach the pod but seems unable to correctly land on it. This is because it cannot correctly handle such a large number of states in the given number of episodes. Also, the environment was discretized to make learning possible.


Episode lengths
Episode length
500 episode average
1000
800
600
400
200
0
2500
5000
7500
10000
12500
15000
17500
20000

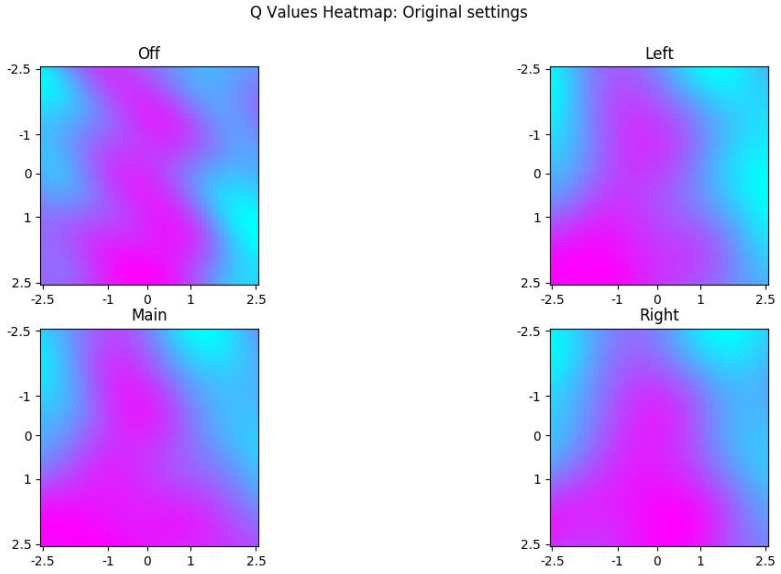
2 Function approximation

Task 3

The implementation was completed but the lander is not able to learn to land properly with the default parameters and number of episodes.



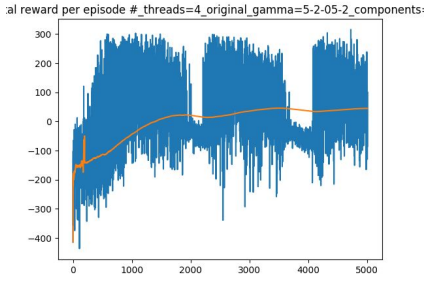
Task 4

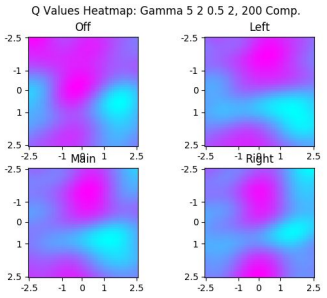


The heatmap shows that the activation values for all the actions are spread around the landing pad position, with higher activations of the engines towards the bottom. The function for off is not very clear, but seems to have learns to be alternatively activated and deactivated in order to smooth the landing.

Question 1

The value of gamma tunes the granularity of the function, with higher values of gamma producing smoother functions. The number of components is related to the precision of the approximation, producing more accurate results when the number is higher. Using the gamma parameters 5, 2, 0.5, 2 the function could better learn to land on the pad. I think this is mainly because of the use of similar values of the left and right engines, that allowed for a proper stabilisation. In addition the model uses 200 components.





Question 2

Linear approximation wouldn't fit the requisites of the task, as it cannot correctly represent the interaction between different features of the state. In this task the interaction is crucial to determine when to turn on and off any engine.